Making Graphical Inferences: A Hierarchical Framework

Raj M. Ratwani (rratwani@gmu.edu)
George Mason University

J. Gregory Trafton (trafton@itd.nrl.navy.mil)
Naval Research Laboratory

Abstract

A hierarchical framework suggesting how graph readers go beyond explicitly represented data to make inferences is presented. According to our hierarchical framework, graph readers use read-offs, integration and pattern extrapolation to make inferences. Verbal protocol data demonstrates high-level differences in the way inferences are made and eye track data examines these processes at the perceptual level.

Introduction

Imagine a scientist examining Figure 1 in order to infer which county in California is going to be hit next by the flu epidemic. How would the scientist go about making this prediction?

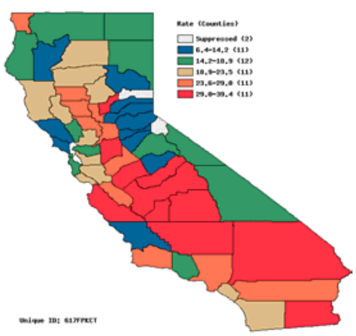


Figure 1. Cases of the flu in California.

Making inferences from graphs is considered one of the more complex skills graph readers should possess. According to the National Council of Teachers of Mathematics (NCTM) the simplest type of question involves the extraction or comparison of a few explicitly represented data points (read-offs) (*NCTM: Standards for Mathematics*, 2003). A more difficult question is an integration question where multiple data points need to be extracted and integrated by some mental operation. The most difficult type of question requires the graph reader to make inferences from the graph. Because the information is not explicitly represented in the data, the graph reader is forced to extrapolate from the current data to make a prediction (Trickett, Ratwani, & Trafton, under review).

How do graph readers go beyond the explicitly represented data to make inferences from graphs? Less is known about how inferences are made from graphical representations, despite the importance of having this skill. Most of the classical theories of graph comprehension (Kosslyn, 1989; Lohse, 1993; Pinker, 1990) do not go into detail about how integration and inferences are made, but instead focus on read-offs from fairly simple graph types. One of the reasons that current theories of graph comprehension do not have much to say about making inferences from graphs is the paucity of data. There are, in fact, very few empirical papers that have systematically investigated how graph readers make inferences from graphs.

We propose a hierarchical framework of graph comprehension for how these different types of questions (read-off, integration, inference) are answered. The most basic type of information extraction is the read-off of explicitly represented data. The more difficult integration of information requires the use of read-off's and spatial transformations (Trafton, Marshall, Mintz, & Trickett, 2002; Trickett & Trafton, in press). For example, in order to integrate information in choropleth graphs (see figures 1 and 2), graph readers read-off specific data points and use spatial transformations by forming clusters of proximate same colored counties and then reason with and compare those clusters (Ratwani, Trafton, & Boehm-Davis, 2003). Finally, in order to make inferences from graphs, we believe graph readers use the same processes used to integrate information (read-off's and spatial transformations) and in addition use pattern extrapolation and mental models (Trafton et al., 2002).

Going beyond the limits of the current data in order to make an inference requires the use of extrapolation (Bott & Heit, 2004); when graph reader's go beyond the limits of visible data, pattern extrapolation may be used. Pattern extrapolation is a process by which graph readers examine known data points and then, based on the pattern of these data points, make an inference.

While the hierarchical framework suggests what cognitive processes graph readers will use when extracting different types of information from graphs, graph readers are likely to use the simplest process to extract the information they desire. For example, when integrating information, if possible, graph readers will use mostly read-offs because read-offs are a simple way of extracting information from graphs and require very little cognitive effort in comparison to spatial transformations or mental model building. Similarly, if a graph reader needs to integrate information from a graph, they are not likely to need to build mental models and extrapolate patterns.

In this paper, we examine which processes are used to make inferences from graphs. We focus on inferences because read-offs are quite well understood (Kosslyn, 1989; Lohse, 1993; Pinker, 1990), there is a preliminary framework for integration of graphical information (Ratwani, Trafton, & Boehm-Davis, 2003), but there are no theories that can adequately describe how inferences are made. Previous research examining graphical inferences has focused on the use of mental models, spatial transformations

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **2004**	2. REPORT TYPE	3. DATES COVERED **00-00-2004 to 00-00-2004**

4. TITLE AND SUBTITLE **Making Graphical Inferences: A Hierarchical Framework**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Research Laboratory,Navy Center for Applied Research in Artificial Intelligence (NCARAI),4555 Overlook Avenue SW,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
In Proceedings of the Twentieth-Sixth Annual Conference of the Cognitive Science Society, 2004

14. ABSTRACT
A hierarchical framework suggesting how graph readers go beyond explicitly represented data to make inferences is presented. According to our hierarchical framework, graph readers use read-offs, integration and pattern extrapolation to make inferences. Verbal protocol data demonstrates high-level differences in the way inferences are made and eye track data examines these processes at the perceptual level.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

and the role of domain knowledge (Hegarty, Shimozawa, & Canham, under review; Trafton et al., 2000). We are interested in how all of these processes are combined in order to make inferences from graphs. In this paper, we focus on read-offs, integration and pattern extrapolation because it is relatively straightforward to identify read-offs, integration, and pattern extrapolation and much more difficult to identify spatial transformations or mental model building.

If the hierarchical framework of graph comprehension is correct, graph readers will make inferences by reading-off explicit information, using pattern extrapolation and integrating information. Experiment 1 serves to explore higher-level thinking about how graph readers make inferences from graphs by using the protocol analysis methodology. Experiment 2 further investigates the processes used to make inferences by using an eye tracker to examine graph readers' eye movements.

Experiment 1

The first experiment was designed to explore the types of processes graph readers use to make inferences from multiple choropleth graphs. Choropleth graphs depicting population densities were selected for use in this experiment; these graphs use different colors, shades of gray, or patterns to represent different quantities. Choropleth graphs were chosen for multiple reasons. First, they are more complex than the graph types used in more traditional studies of graph comprehension and better reflect how graphs are used in the real-world. Second, these particular graphs do not require a great deal of domain knowledge and can be presented to undergraduates without much training. Finally, choropleth graphs represent a class of graphs that are commonly used by scientists in such domains as meteorology, geology and oceanography.

Method

Participants

Three George Mason University undergraduate psychology students served as participants for course credit. Informed consent was received from all participants.

Materials

Twenty sets of choropleth graphs were created; each set consisted of three conceptually related graphs. The graphs in each set displayed the population densities of fifty fictitious counties. The first graph in each set displayed the population from the year 1990, the second graph displayed 1995 and the third graph displayed the population from the year 2000 (See Figure 2 for an example). Only one county in each set of graphs was labeled with a county name (referred to as the target county) in order to reduce search time. Previous studies found that graph readers spent a great deal of time searching for the county of interest when every county was labeled (Ratwani et al 2003). One inference question was asked of each set of graphs: What will the population of the target county be in the year 2005?

Design

Five sets of graphs showed a clear decrease in the overall population densities from 1990 to 2000 while the population of the target county did not change in any of these graphs. These counties surrounding the target county had a powerful contextual indication that the population was decreasing. Five sets of graphs showed a clear increase in the overall population densities from 1990 to 2000 while the population of the target county did not change. These counties surrounding the target county had a powerful contextual indication that the population was increasing. Ten of the sets of graphs served as fillers and were removed from all analyses; the populations were jumbled and had increasing, decreasing or no clear pattern to the population movement across the graphs. The purpose of these sets was to randomize the patterns of increasing and decreasing population in the ten sets of interest. The order in which the twenty sets of graphs were presented was randomized for each participant. The increasing and decreasing sets were combined in the analyses below.

Procedure

All participants first read the question for the set of graphs they were about to view and then examined the graphs. For example, the participant would read the question, "What will the population of county x be in the year 2005?" After reading the question the participant would then view the graphs from the three time periods and make their prediction. This process continued for each of the twenty sets of graphs.

The participants could view each of the graphs for as long as they desired, and the participants were permitted to look back to any of the graphs within a particular set as needed. Each graph was presented on a single sheet of paper. After answering each question, the participant went on to the next set of graphs. Each participant provided a talk-aloud protocol (Ericsson & Simon, 1993) as they examined the graphs and answered the questions. The participants' verbal protocols and the graphs they were examining were videotaped.

Coding Scheme

Transcriptions of the verbal protocols were made prior to data analysis. The first step was to segment the protocols into individual utterances. Utterances were defined as a single thought and utterances that were not germane to the task at hand were coded as "off task" and eliminated from further analysis. Each remaining utterance was then coded according to our hierarchical framework. The utterances were coded as either being a **target read-off** (extracting information regarding the target county of interest only) or **integrative** (extracting general trend information from the graph). All the answers were, by definition, inferences. There were no non-target read-off's in any of the utterances made by the participants. Table 1 shows examples of each utterance type.

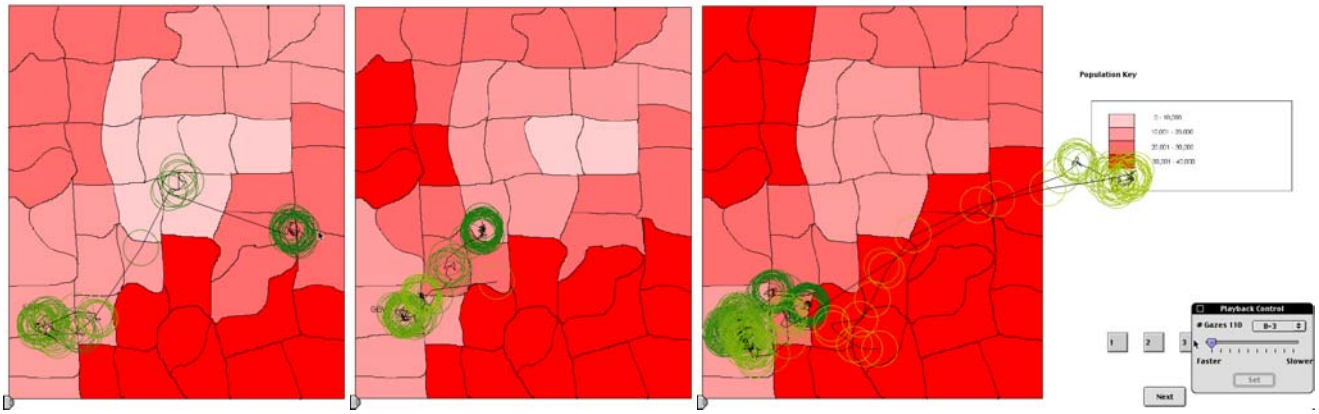


Figure 2. Graphs depicting population growth from 1990, 1995, and 2000 (left to right).

Code	Example
Target read-off	In 1990 Stow County was 20,000 to 30,000
Integrative	All the other areas are increasing in population

Table 1: Examples of each extraction type.

Results and Discussion

Our main goal was to explore how graph readers made inferences from graphs. We first examined the types of extractions made by each graph reader and then compared these extractions to the answer given to the inference questions. The participants gave a numerical answer indicating that the population of the target county would either change or not change. Of the three participants, one participant made change responses the majority of the time, one participant made non-change responses the majority of the time and one participant had mixed responses. Thus, graph readers were not always using the same strategy to make these inferences. When participants made a change response, their inference was in the direction consistent with the surrounding counties. For example, when the participant made a change response and inferred that the population of the target county would grow in the future, the surrounding counties were also growing.

As figure 3 suggests, when graph readers said the population of the target county would not change in the future all of their extractions were target read-offs. When graph readers said that the target county would change in the future, the graph readers made some target read-offs but made a significantly greater number of integrative extractions, $\chi^2(1) = 4.9$, $p < .05$. In addition, when a non-change response was given, graph readers made a significantly greater number of target read-offs than when a change response was made, $\chi^2(1) = 8.02$, $p < .01$.

The verbal protocol data indicates that when graph readers performed mostly target read-off's they made a non-change response to the inference questions. That is, despite the fact that the powerful overall context of the three graphs suggested that the population was increasing (for example), the graph reader inferred that the population of the target county would not change in the future. However, when graph readers looked beyond the target and used the global context of the graphs the graph readers used context to infer that the growth would continue to the target county and that the target county would change in the future.

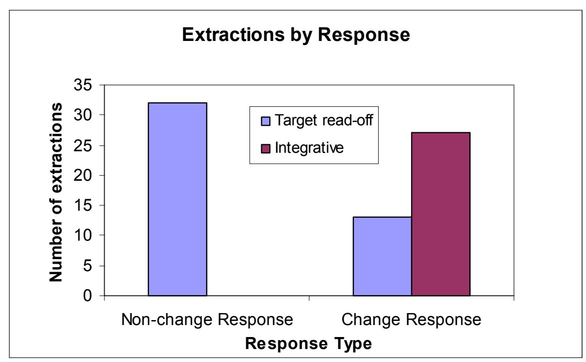


Figure 3: Number of extractions by response.

These data suggest there are differences in the way people think about making inferences. Based on the type of extractions the graph reader made, their response could be categorized as either inferring a change or not inferring a change in the future population. There appear to be two general ways in which graph readers made inferences from these graphs. One way was to focus only on the target county in each of the three time periods and, based on how the population changed in the target county, an inference was made as to the future population. For example, if the target county did not change population in any of the three time periods then it would not change in the future. This no change response appears to be based solely on pattern extrapolation of the target county. Alternatively, when making change responses, graph readers appeared to be making read-off's for pattern extrapolation and taking into

consideration the contextual influences of the graph. Some aspects of the global context of the graph were being integrated with the population of the target county in order to make the inference. When participants made change responses, their verbal protocol data is consistent with an interpretation of them creating a dynamic mental model: participants were imagining the growth in the counties extending to nearby counties, eventually "hitting" the target county. While we are not directly measuring mental model formation in this paper, we are interested in what information is needed to form those mental models.

Based on our hierarchical framework, we would expect graph readers to make inferences by both reading-off information for pattern extrapolation and integrating information. It appeared that when graph readers made a non-change response, they extracted target information only, noticed it did not change, and extrapolated that it would not change in the future. They did not seem to explicitly extract information from nearby counties. Graph readers who made a change response appeared to be using read-offs, pattern extrapolation and integration as our hierarchical framework suggests.

Experiment 1 showed that people had different strategies when answering inference questions: a change strategy and a non-change strategy. It could be that at the perceptual level these strategies are identical. For example, it could be that participants who made a no-change answer did, in fact, look all over the graph but decided to simply ignore that information, or assume that the target county was the most important determinant of future change. Additionally, the protocol data did not show how or what types of information was extracted by change-response participants. Experiment 2 investigates these issues.

Experiment 2

How, then, did participants make inferences from these graphs? By performing a small task analysis, it is obvious that when information needs to be integrated, it can be integrated in at least two ways: within a specific graph and between related graphs. If a participant integrates information within a specific graph, the participant would presumably examine nearby counties to see how their population was different from the target county. If a participant integrates information between related graphs, the participant would probably examine graphs that had changed over time.

Experiment 2 will explore three main issues. First, do participants who answer change and non-change have different perceptual strategies? Second, what types of integration do change participants engage in (within graphs, between graphs, or both)? Third, what is the proportion of read-offs and integration used in order to answer inference questions and how do those proportions relate to the answers that participants gave?

Method

Participants

Thirteen George Mason University undergraduate psychology students served as participants for course credit. Informed consent was received from all participants.

Materials

The same sets of graphs used in the first experiment were used in the second experiment. In this experiment the materials (graphs and questions) were displayed on a computer screen. Eye track data was collected using an LC Technologies Eye gaze System eye tracker operating at 60Hz (16.7 samples/second).

Design

The design was the same as Experiment 1.

Procedure

The procedure was very similar to that used in Experiment 1; however, the use of the computer and eye tracker did necessitate some changes. The participants were seated at a comfortable distance from the monitor and used a chin rest. Participants first were calibrated on the eye tracker. Participants were then shown the question at the top of a blank screen and read the question out loud. Previous studies (Ratwani et al., 2003) have shown that the process of collecting eye track data was not hindered by the participant talking. After reading the question the participant proceeded to the first graph. The interface allowed the participants to progress from graph to graph within a set with a button-click. The participants were instructed to say their answer out loud when they made their inference. After answering the question, the participant could progress to the next question and set of graphs.

Coding Scheme

A gaze was defined from each sample being no more that 10 pixels in Euclidian distance from the center of gravity of the previous point for at least 100 milliseconds. Frequencies were created by counting the number of gazes to different areas of the graph. The areas of the graph that were coded were: the legend, the title of the graph, and the main part of the graph itself.

Participant's gazes to the main part of the graph were coded to examine how much reading-off and how much integrating the participants were doing. Gazes to the target county were coded to examine how often participants were making read-offs. Gazes to locations other than the target were coded in order to examine whether graph readers were integrating information within the graph. In order to capture how far away from the target county participants were gazing the location of the gaze relative to the target county was coded for. For example, if a participant gazed at a county that was three counties away from the target this distance was coded.

Integration of information between graphs was coded by examining areas of change relative to the previously viewed

graphs. If a gaze to the graph was to a county where the population value changed from a previously viewed graph this was coded for. For example, if the participant gazed to a county in 1995 that had changed in population relative to the map from 1990 then this was coded as a change gaze. Thus, the first map viewed by each participant in every set did not have any change gazes.

Results and Discussion

Experiment 2 was designed to examine what processes occurred at the perceptual level when graph readers made inferences from graphs. Specifically, we wanted to further investigate the process differences when graph readers made a change response as compared to when they made a non-change response.

The responses made by graph readers were mixed, but mostly (76%) change responses were made. The raw frequencies of gazes were normalized by dividing the frequency of gazes by the number of responses in either the non-change or change category.

There were no significant differences in the number of target gazes when participants made a non-change response as compared to a change response as figure 4 suggests, $\chi^2(1) = .34$, $p = .56$. Participants appeared to be reading off the same amount of target information regardless of what type of response they made. Thus, participants were reading off information nearly equally when making inferences.

However, participants who made change responses did more integration within the graph than participants who did not make change responses. As figure 4 suggests, those who made change responses on average made a greater number of gazes to counties other than the target, $\chi^2(1) = 4.52$, $p < .05$.

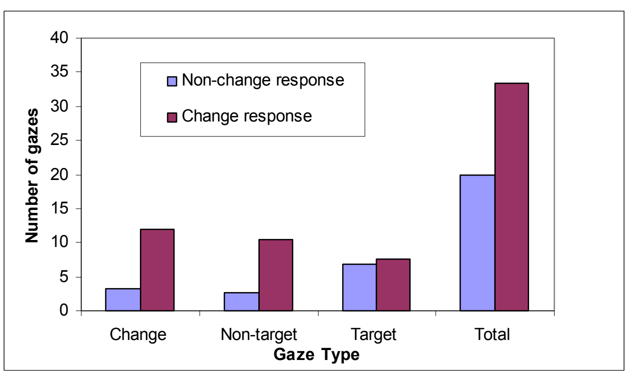
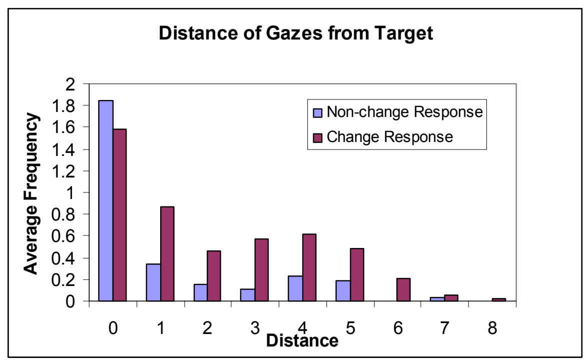


Figure 4. Number of coded gazes.

How far away from the target did participants look? In order to examine this issue, we created histograms showing the location of the counties that were gazed at based on the participants response. Figures 5 shows the frequency of gazes participants made to the target and to counties other that the target. The x-axis shows how far away the county gazed at was from the target. Zero represents the target and one through eight represent how far away the county gazed at was from the target. The patterns in these histograms are significantly different, $\chi^2(8) = 22.82$, $p < .01$, suggesting that when graph readers made a change response, they frequently looked at the target and counties away from the target, whereas graph readers who made a non-change response focused primarily on the target. Consistent with this interpretation, the proportion of change gazes to non-targets (68%) was far greater than the proportion of non-change gazes to non-targets (37%), $\chi^2(1) = 9.2$, $p < .005$. Graph readers who made a change response were frequently looking at counties as far as 6 away from the target county.

Did graph readers integrate information between graphs when they were making inferences? Integration between graphs was examined by looking at the number of gazes to areas of change from one graph relative to another. As figure 4 suggests, participants who made change responses made a significantly greater number of gazes to areas of change as compared to participants who made non-change responses, $\chi^2(1) = 4.88$, $p < .05$. This suggests participants who made change responses were integrating information between graphs by comparing the areas that changed in population.



Figures 5. Histogram of distance by response.

The process of integrating information between graphs is further supported be examining the number of times graph readers examined each of the three graphs. For example, some participants viewed each graph once; the sequence of graphs they looked at was 1→2→3. Whereas other graph readers examined each graph more than once and had a sequence such as 1→2→1→2→3→2→3. Participants who made change responses looked at the three graphs in each set more often in order to compare the counties that changed population between graphs, $\chi^2(2) = 5.24$, $p < .05$. Thus, graph readers who made change responses integrated information between graphs by paying attention to areas that changed in the graphs and looking at the graphs frequently in order to make these comparisons.

When participants made non-change responses, their eye movements suggest they are primarily examining the target county. These participants generally looked at each map only once. Furthermore, they appeared to be reading-off target county information in each graph, noticing the pattern does not change, and then using pattern extrapolation to infer that the pattern will not change in the future.

Graph readers making change responses appeared to read off target county information and also focused a great deal of attention on non-target counties. These gazes to non-target counties appeared to be a way to integrate the information from the other counties with the information about the target county. These graph readers also integrated information between graphs by paying attention to areas of change between the graphs. Finally, they compared areas of change to infer the future population by looking back and forth at the graphs in each set. This is suggestive evidence of the formation of dynamic mental models which may be used to understand how the contextual growth or decay is influencing the target county.

General Discussion

How do people make inferences from graphs? Most classic theories do not provide any mechanisms for making graphical inferences. These studies examined inferences at a high-level by focusing on graph reader's thought processes with the verbal protocol data and also at the perceptual level by examining graph reader's eye movements. These studies demonstrate that people certainly can make inferences, and that people make inferences in different ways. One way that people make inferences is to examine the specific object that will change over time (target county in our case). Depending on the type of change that is observed, a pattern is extracted and then extrapolated. In our studies, approximately a quarter of the answers conformed to this strategy. The remainder took context into account. That is, they observed the surrounding counties (especially the ones that changed from graph to graph) and presumably imagined the change affecting the target county.

Our hierarchical framework of graph comprehension is consistent with both views, though it is supported more strongly by the participants who made change answers. In order to make inferences, the hierarchical view framework suggests that people need to extract specific information from graphs (well described by most theories of graph comprehension), integrate information into a reasonable whole (in this case by combining information between and within graphs), use that information to extrapolate beyond the given data, mentally manipulate the graphical information by spatial transformations and build mental models. It is interesting that when non-change answers were made, only a subset of this framework was used: the evidence for integration in particular was quite weak. It seems that when non-change answers were given, participants simply took in the specific information for the target county, performed simple extrapolation, and then gave an answer. Using the surrounding counties was just not a priority for these participants.

Finally, how inferences are made from graphs is more complex than we have described here. Our hierarchical framework identifies the processes used to make inferences; however, further empirical data is needed to understand how read-offs, integration, spatial transformations, mental models and pattern extrapolation are combined in the process of making inferences. In addition the processes outlined by our hierarchal framework are likely to be dependent on many factors such as knowledge of the graphical display and domain knowledge (Hegarty et al., under review).

Acknowledgements

This research was supported in part by grant 55-7850-00 to the second author from the Office of Naval Research and by George Mason University. We thank Mike Schoelles for designing the interface of the eye tracker.

References

Bott, L., & Heit, E. (2004). Nonmonotonic Extrapolation in Function Learning. Journal of Experimental Psychology: Learning, Memory & Cognition, 30(1), 38-50.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analysis: Verbal reports as data. (2nd ed.). Cambridge, MA: MIT Press.

Hegarty, M., Shimozawa, N., & Canham, M. (under review). Inferences from Graphics: Do you need a Weatherman to Tell which Way the Wind Blows?

Kosslyn, S. M. (1989). Understanding charts and graphs. Applied Cognitive Psychology, 3, 185-225.

Lohse, G. L. (1993). A cognitive model for understanding graphical perception. Human Computer Interaction, 8, 353-388.

NCTM: Standards for Mathematics. (2003, 2003).

Pinker, S. (1990). A theory of graph comprehension. In R. Freedle (Ed.), Artificial intelligence and the future of testing (pp. 73-126). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ratwani, R. M., Trafton, J. G., & Boehm-Davis, D. A. (2003). Thinking graphically: Extracting local and global information. In R. Alterman & D. Kirsch (Eds.), 25th. Annual Meeting of the Cognitive Science Society. Boston, MA: Erlbaum.

Trafton, J. G., Kirschenbaum, S. S., Tsui, T. L., Miyamoto, R. T., Ballas, J. A., & Raymond, P. D. (2000). Turning pictures into numbers: Extracting and generating information from complex visualizations. International Journal of Human Computer Studies, 53(5), 827-850.

Trafton, J. G., Marshall, S., Mintz, F., & Trickett, S. B. (2002). Extracting explicit and implicit information from complex visualizations. In M. Hegarty, B. Meyer, & H. Narayanan (Eds.), Diagrammatic Representation and Inference (pp. 206-220). Berlin Heidelberg: Springer-Verlag.

Trickett, S. B., & Trafton, J. G. (in press). Spatial Transformations in Graph Comprehension. Diagrammatic Representation and Inference, 3rd International Conference, Cambridge, U.K.

Trickett, S. B., Ratwani, R. M., & Trafton, J. G. (under review). Real World Graph Comprehension: High-Level Questions, Complex Graphs, and Spatial Cognition.